



oncrawl



Oncrawl x ChatGPT Webinar Part I: Tracking ChatGPT indexation and visibility with Oncrawl's log analysis

Speakers



Janaina Barreto-Romero

Senior Technical SEO
@Oncrawl



Jérôme Salomon

Senior Technical SEO
@Oncrawl



Before we start

- **Welcome!**
- There will be a **replay** – you'll receive it by email in a couple days.
- Don't hesitate to **ask questions** in the **dedicated tab** at the bottom of your screen.
- We will answer them at the end of the presentation.



Agenda

- Global search context
- How ChatGPT Search works
- Optimizing for ChatGPT Search – is it possible?
- OpenAI crawlers
- Indexation & visibility in ChatGPT Search
- Using Oncrawl's log analysis
- Key takeaways
- What's next?
- Q&A

Global Search Context

Search engines market share

Worldwide market share of search engines (Desktop) – January 2025

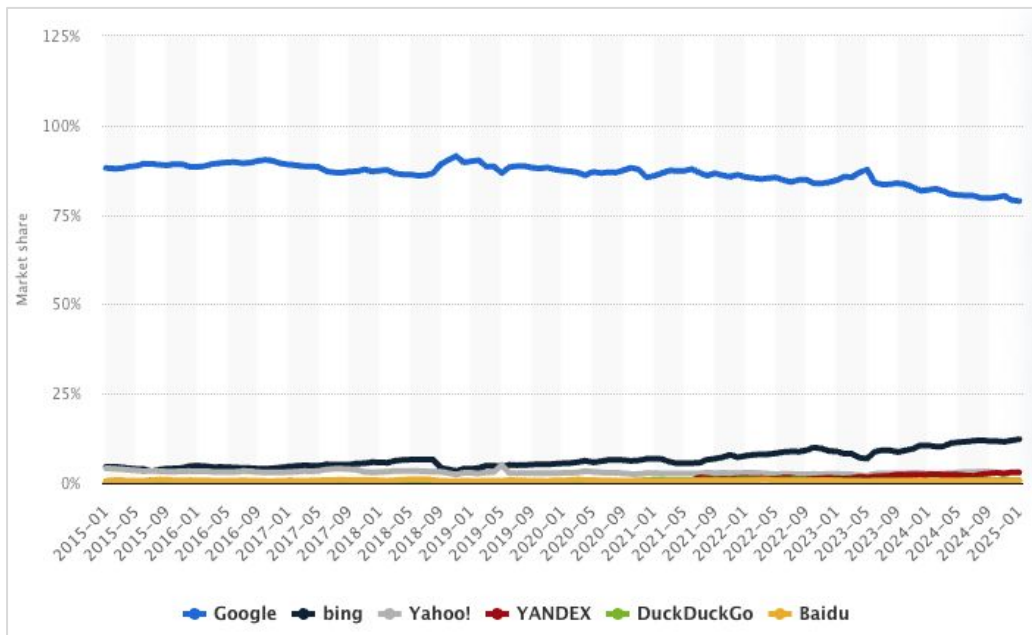
- Google: 78.83 %
- Bing: 12.23 %
- Yahoo: 3.07 %

Google largely dominates, but some engines remain locally influential:

- Yandex in Russia
- Baidu in China
- Yahoo in Japan and Mexico

Search engines market share

But in the last 10 years we see that Bing has progressed against Google: +7.8pts



2015-01	
Google	88.1%
bing	4.53%
Yahoo!	4.13%
YANDEX	0.35%
DuckDuckGo	0.05%
Baidu	0.61%

2025-01	
Google	78.83%
bing	12.23%
Yahoo!	3.07%
YANDEX	2.94%
DuckDuckGo	0.87%
Baidu	0.77%

Source : [statista.com](https://www.statista.com)

ChatGPT leading the surge in AI search engines

Jul 2024: Launch of SearchGPT prototype

Oct 2024: SearchGPT is included in ChatGPT and becomes **ChatGPT Search**

Nov 2024: +44% monthly growth in referring traffic from chatgpt.com

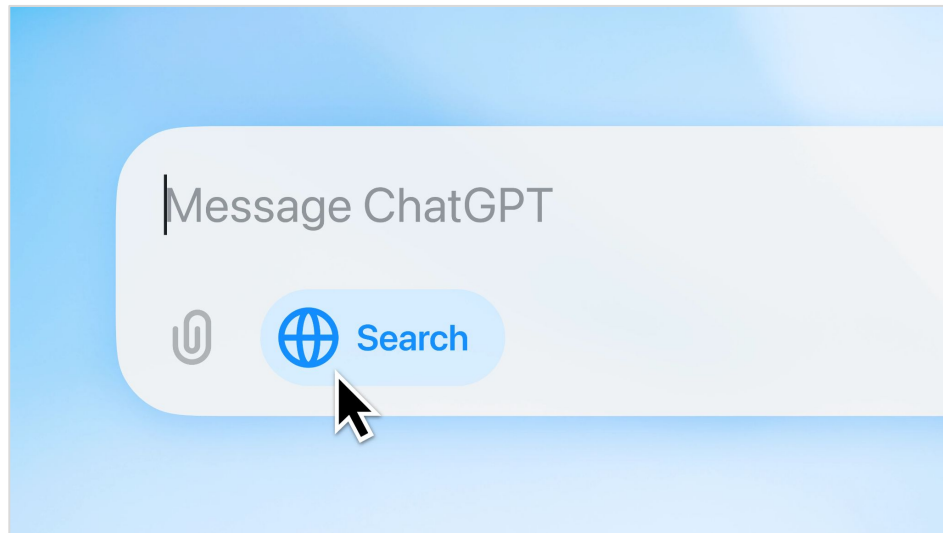
Dec 2024: ChatGPT generates 6 times more clicks than Perplexity

ChatGPT is on a trajectory to potentially capture a 1% market share in 2025.

AI search engines are disrupting usage, but Google remains dominant.

What is ChatGPT Search?

ChatGPT Search activation






ChatGPT Search feature is **manually initiated by the user**



Or based on the prompt:


- If the user asks for a source or a link.
- To provide local or updated results.
- To point to official documents, tools, APIs, etc.

ChatGPT Search answers

what are some great ways to fix up a backyard


  






 **54 Backyard Ideas to Upgrade Your Outdoor Space**
thespruce.com > backyard ideas 



Transforming your backyard into a welcoming and functional space can be both enjoyable and rewarding. Here are several ideas to enhance your outdoor area:

- 1. Create a Cozy Seating Area:** Arrange comfortable outdoor furniture, such as sofas, chairs, or hammocks, to establish a relaxing spot for reading or entertaining guests. Adding colorful cushions and throw pillows can infuse personality and comfort. THE SPRUCE
- 2. Incorporate Outdoor Lighting:** Enhance ambiance and extend usability into the evenings by installing string lights, lanterns, or solar-powered fixtures. Soft, warm lighting can create a magical atmosphere. THE SPRUCE
- 3. Add a Fire Pit:** Introducing a fire pit provides a focal point for gatherings and allows for enjoyable

Citations 

-  The Spruce
43 Backyard Ideas on a Budget — The Spruce
August 25, 2024 — Set up a hammock and add pillows, string lighting in trees, blankets, and even a plate (or b...
-  The Spruce
7 Ways to Refresh Your Backyard This Summer Without Spending a Dime — The Spruce
May 18, 2024 — Power wash the patio. Clean outdoor rugs. Wipe down furniture. Wash cushion covers and...
-  Family Handyman
50 Brilliant Ways to Spruce Up Your Backyard This Summer
October 6, 2024 — Make Your Own Cornhole Game. Cornhole boards are a classic backyard game that is e...
-  Better Homes & Gardens
50 Stunning Backyard Ideas That Fit Every Kind of Space...
May 7, 2024 — Whether you're looking for patio suggestions, outdoor dining, or a fun space to relax wi...
-  Extra Space Storage
Your DIY Guide to a Backyard Makeover on a Budget
September 24, 2024 — Give your outdoor living space a facelift with a calming meditation space. Find a seclud...

- ChatGPT Search provides fast, timely **answers including links to relevant web sources.**
- It blends natural language with **up-to-date information with real-time search on third party providers.**
- Aims for better answers by considering chat context and adding **citations to the content source.**

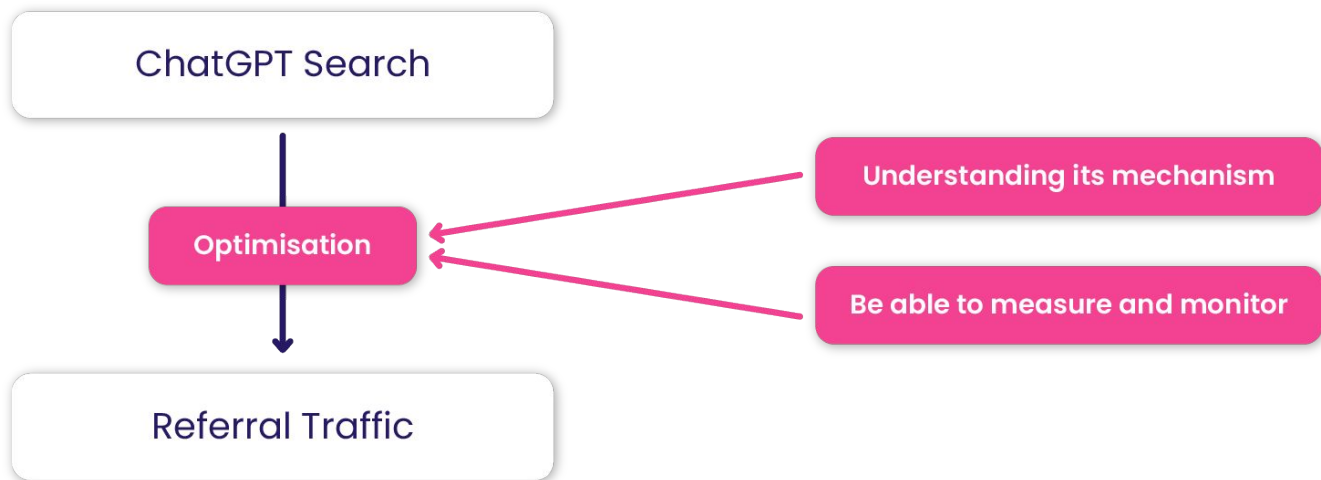
Optimizing for ChatGPT Search

Optimizing for ChatGPT Search

ChatGPT Search is a **growing source of referral traffic**

You can optimize for this new opportunity if you:

- Understand how ChatGPT Search works
- Are able to measure and monitor your efforts



The challenge of reliable data

Measuring and monitoring your SEO efforts on ChatGPT Search is challenging:

- **There's no equivalent of Google Search Console for ChatGPT Search**



- **Traditional SEO metrics are not available or meaningless:**
 - No keyword search volume
 - No average position, impressions, clicks and CTR
 - No state of indexation

What are SEO's using to measure so far ?

Tracking referral traffic



- Analytics tools
- Not enough to measure your efforts

Scraping ChatGPT Search Answers



- Visibility tracking tools
- Users' queries and search volume remain only assumptions

Why SEO should start using log analysis ?

OpenAI is using different **crawlers to run ChatGPT Search**.

While crawling your website, these bots are leaving valuable clues in your server log files:

- **Indexation: Verify that ChatGPT has access to your content**
- **Visibility: Use this bot traffic as a visibility metric**

Log analysis might be the **best source of insights for indexation and visibility tracking of your website in ChatGPT Search**.

OpenAI crawlers

Meet OpenAI's crawlers

OpenAI is using different bots/crawlers for different tasks and features:

ChatGPT "offline"

GPTBot



Crawling webpages for model training.

ChatGPT Search

ChatGPT-User



Crawling webpages to answer a user query in real-time.

OAI-SearchBot



Crawling webpages to improve Search features asynchronously.

OpenAI's bots and robots.txt

You can allow or disallow OpenAI's crawlers case-by-case.

To have your website **appear in ChatGPT Search results**, you have to **allow both OAI-Searchbot and ChatGPT-user** in the robots.txt.

```
#Allow AI Search  
ChatGPT-User:  
OAI-Searchbot:  
Allow: /
```

You can **disallow OpenAI's training bot** to prevent ChatGPT's "offline" models from being trained on your content **but still appear in ChatGPT Search results**.

```
#Disallow AI Training  
ChatGPT:  
Disallow: /
```

OpenAI's AI search crawlers

ChatGPT-user and **OAI-SearchBot** are the key bots to monitor.

ChatGPT Search system relies on them:

- **ChatGPT-user** crawls webpages to answer a user in real-time.
- **OAI-SearchBot** crawls webpages to "index" and improve the search process.

ChatGPT-User



**Useful to track
visibility**

OAI-SearchBot



**Useful to track
indexation**

How log analysis works

Every time a bot or a user visits a page on your website, it leaves a new line in your server log files.

Each line contains the following information used for log analysis in SEO:

- URL visited
- Date and time
- User-agent (to identify the botname)
- Referrer (to identify the source of the user visit)
- Status code of the URL

How to identify OpenAI's crawlers

GPTBot user-agent: *Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko); compatible; GPTBot/1.1; +https://openai.com/gptbot*

```
20.171.207.154 fr.oncrawl.com - [02/Apr/2025:02:04:27 +0000] "GET /2017/webinar-digest-optimiser-budget-crawl/ HTTP/1.0" 301 162 "-" "Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko; compatible; GPTBot/1.2; +https://openai.com/gptbot)"
```

OAI-SearchBot user-agent: *OAI-SearchBot/1.0; +https://openai.com/searchbot*

```
51.8.102.102 fr.oncrawl.com - [02/Apr/2025:01:04:47 +0000] "GET /referencement/seo-scotiabank-communication-covid-creation-hub-educatif/ HTTP/1.0" 200 71566 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/131.0.0.0 Safari/537.36; compatible; OAI-SearchBot/1.0; +https://openai.com/searchbot"
```

ChatGPT-user user-agent: *Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko); compatible; GPTBot/1.1; +https://openai.com/gptbot*

```
52.156.77.152 www.oncrawl.com - [02/Apr/2025:01:12:06 +0000] "GET /technical-seo/how-does-a-browser-create-a-web-page/ HTTP/1.0" 200 84974 "-" "Mozilla/5.0 AppleWebKit/537.36 (KHTML, like Gecko); compatible; ChatGPT-User/1.0; +https://openai.com/bot"
```

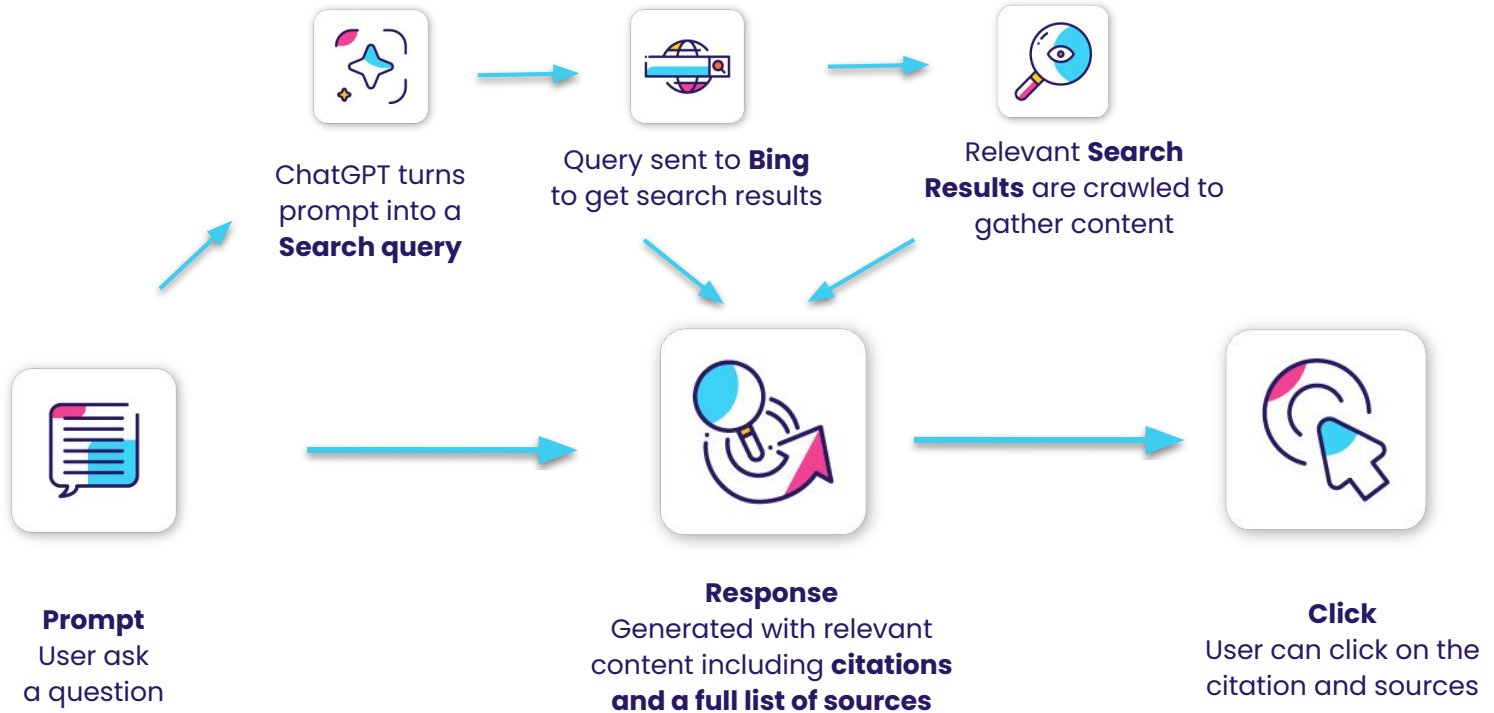
ChatGPT clicks / referral traffic

- You can also use server logs to track referral traffic from ChatGPT
- When a user clicks on a link on chatgpt.com pointing to your website, your server log will register a new line with :
 - Visited URL
 - Status code
 - Date and time
 - Referer value: "https://chatgpt.com" (or "https://chat.openai.com")

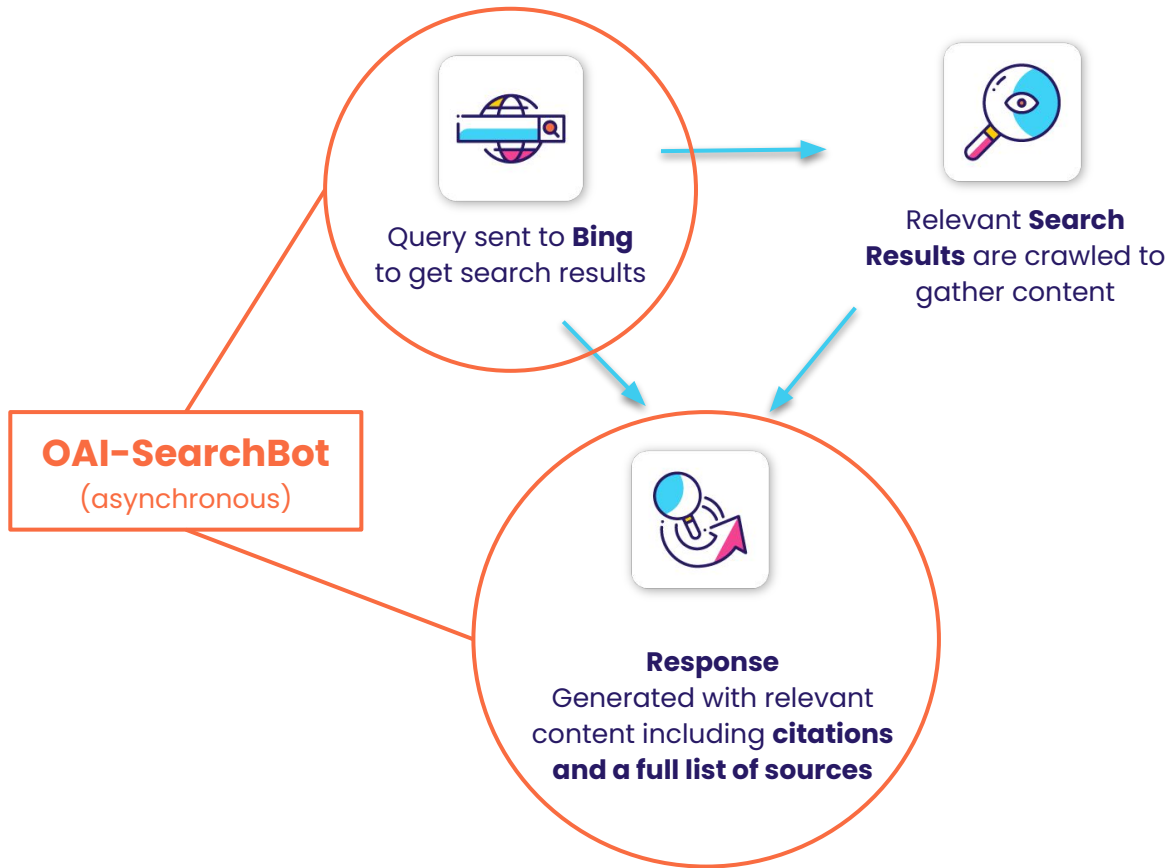
```
██████████ www.oncrawl.com - [03/Apr/2025:11:33:02 +0000] "GET / HTTP/1.0" 200 47626  
"https://chatgpt.com/" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/  
537.36 (KHTML, like Gecko) Chrome/134.0.0.0 Safari/537.36"
```

Indexation and visibility in ChatGPT Search

From prompt to click



OpenAI's bots role in the search process



The real-time search process

- ChatGPT turns the user prompt into a search query, taking into account the conversation and user context.
- ChatGPT performs a real-time search using this query and relies on third-party search engines: Bing, news partners, and others search engines.
- ChatGPT will surface the search results (a list of URLs) to identify relevant content.
- **OAI-SearchBot complements this process** by ensuring that relevant content is accessible and usable within OpenAI's ecosystem.

OAI-SearchBot's crawling process

- OAI-SearchBot crawling is asynchronous to user queries
- OAI-SearchBot mainly relies on **Bing's search index and SERPs to locate relevant pages during web browsing.**
- If your site is indexed in Bing, it's much more likely to appear in those results and be visited by the bot.
- If your site is not indexed in Bing, it may still be reachable if the exact URL is provided or **linked from other indexed sources**, but discoverability will be limited.

OAI-SearchBot's indexing process

- **OAI-SearchBot is ChatGPT Search's "indexing" bot.**
- While OAI-SearchBot collects and processes data, OpenAI does not maintain a traditional search index like Google or Bing.
- The data collected is used to improve:
 - How ChatGPT interacts with third-party search engines and processes retrieved content.
 - The relevance of search results.
- **OAI-SearchBot's "indexing" has an impact on the search results ranking.**

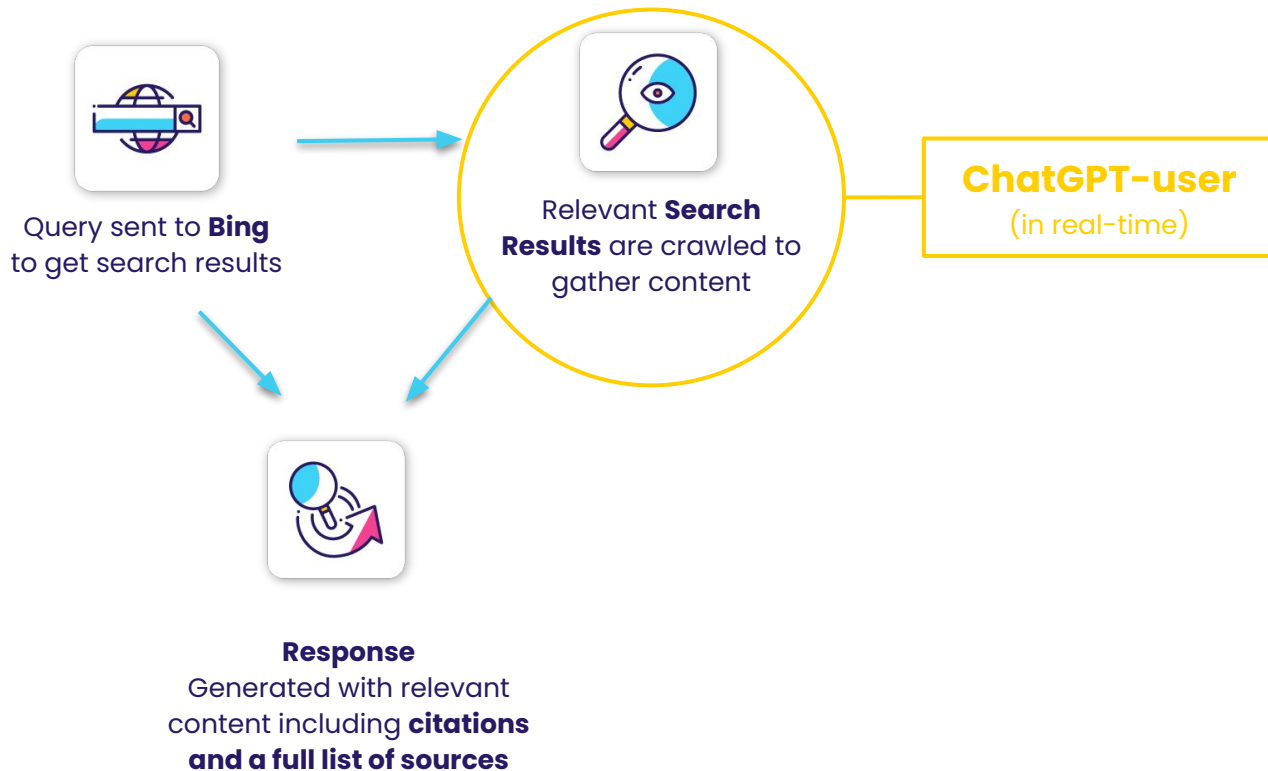
OAI-SearchBot's indexing process

- **OAI-SearchBot does not “index” all visited pages**
- **Traditional SEO signals are considered to define if a page is eligible:**
 - Robots.txt
 - Status code
 - Meta robots (noindex)
- OpenAI's content policies are also applied as filters

The value of OAI-SearchBot bot hits

- You can track this bot to prevent indexation issues.
 - Ensure your key content is accessible and crawled
 - Look for status code and server response time
- Even if **OAI-Searchbot** doesn't operate at the time of user interaction, its visits to your website are a signal of **ChatGPT's interest** in your content:
 - Monitor bot volume over time trend
 - Ensure key content is regularly crawled
 - Identify most crawled pages and content type
- You can also use log analysis to check the **last time OAI-Searchbot crawled a page**, ensuring it picked up your updates.

OpenAI's bots role in the search process



ChatGPT-user and search results

- **ChatGPT-user** is dispatched to crawl pages found in the search results to fetch relevant information.
- **But ChatGPT-user does not crawl all search results:**
 - Webpage protected by a paywall
 - Blocked by robots.txt
 - Inaccessible due to server restrictions
- Search results pages not crawled by ChatGPT-user will still be listed in the source tab.
- **Not all pages crawled by ChatGPT-user will be used as a citation** in the answer.

ChatGPT-user and caching

Caching mechanism is in place for efficiency, which may affect ChatGPT-User behavior:

- If a page was recently accessed by ChatGPT-User and remains relevant, it may not be re-crawled immediately.
- However, for content that is dynamic or time-sensitive, the bot may revisit the page to ensure the response includes the latest information.
- If OAI-SearchBot has previously processed and contributed to search improvements for a particular site or topic, the enhanced search model may reduce the need for ChatGPT-User to crawl that same page.

The value of ChatGPT–User bot hits

Tracking ChatGPT–User’s crawl activity on your website is by far the **most interesting metric**:

- If ChatGPT–user visits a page, it means ChatGPT Search is accessing your content in real-time following a user query.
- It allows you to measure the **“volume of prompts” that require the use of your content.**
- To date, the **volume of bot hits from ChatGPT–user** might be the **most reliable visibility indicator for ChatGPT Search.**
- You may use it as an equivalent to impressions on Google.

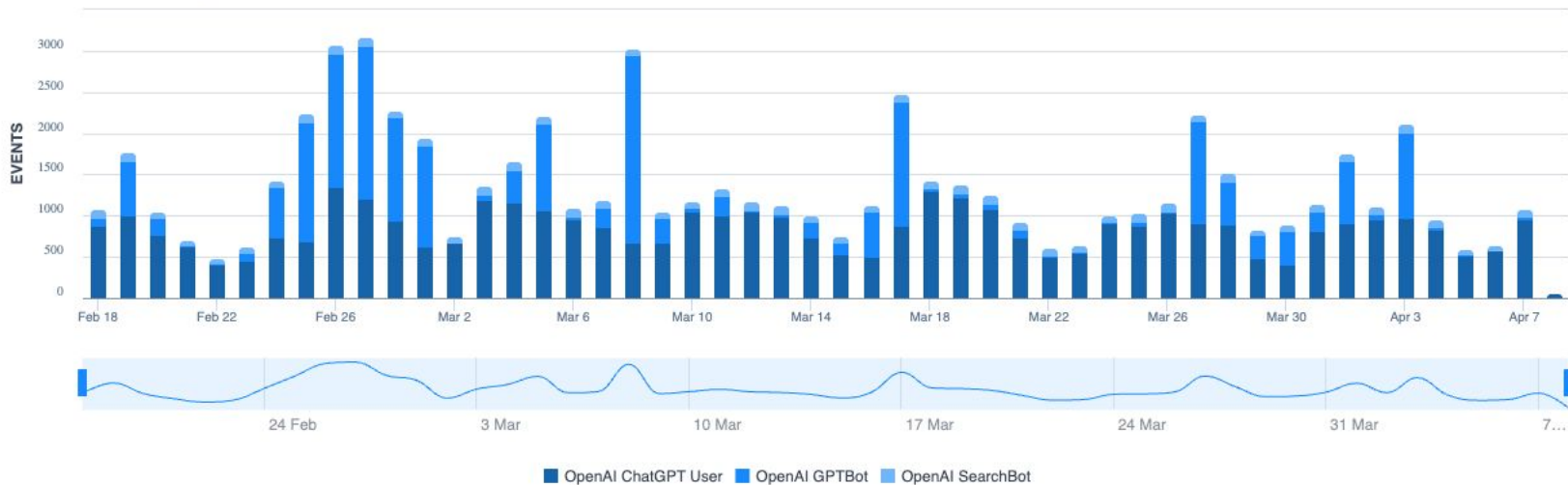
Log analysis with Oncrawl

Metrics and filters available

- **Key metrics**
 - Volume of bot hits
 - Volume of unique pages crawled
 - Volume of referral traffic
 - Volume of unique pages with referral traffic
- **Filters**
 - Bot name
 - Status code
 - Segmentation (type of pages)
- **Timeseries**
 - Total of selected period of time
 - Day by day evolution
 - Hour by hour evolution

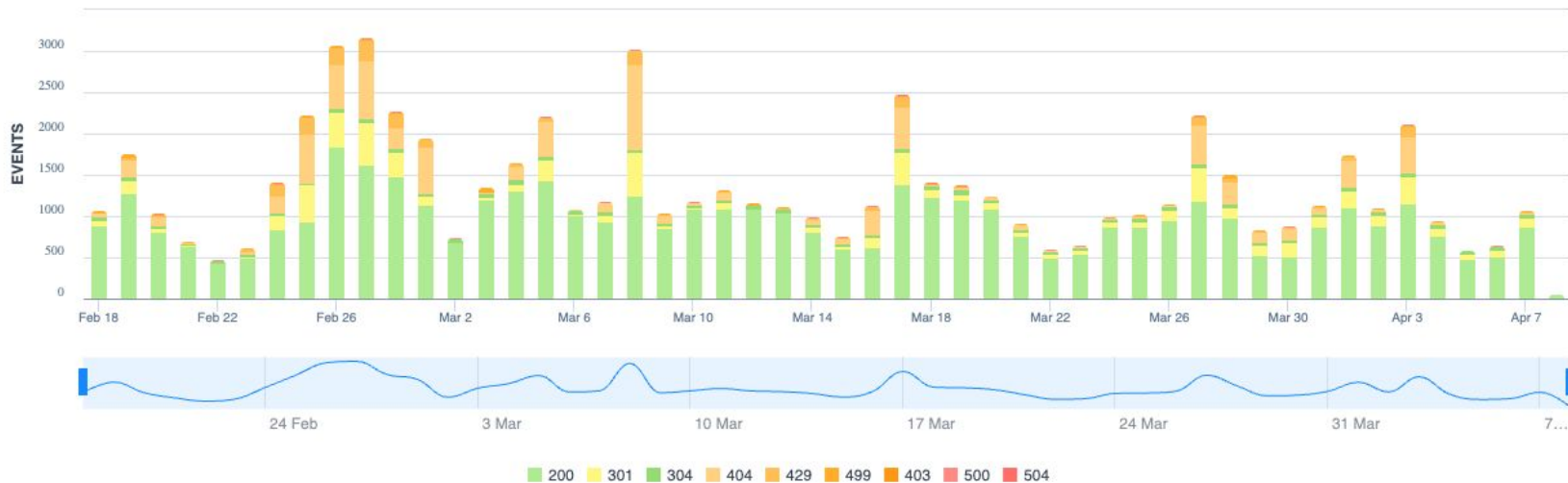
Compare the OpenAI's bots behavior

Bot hits by user agents



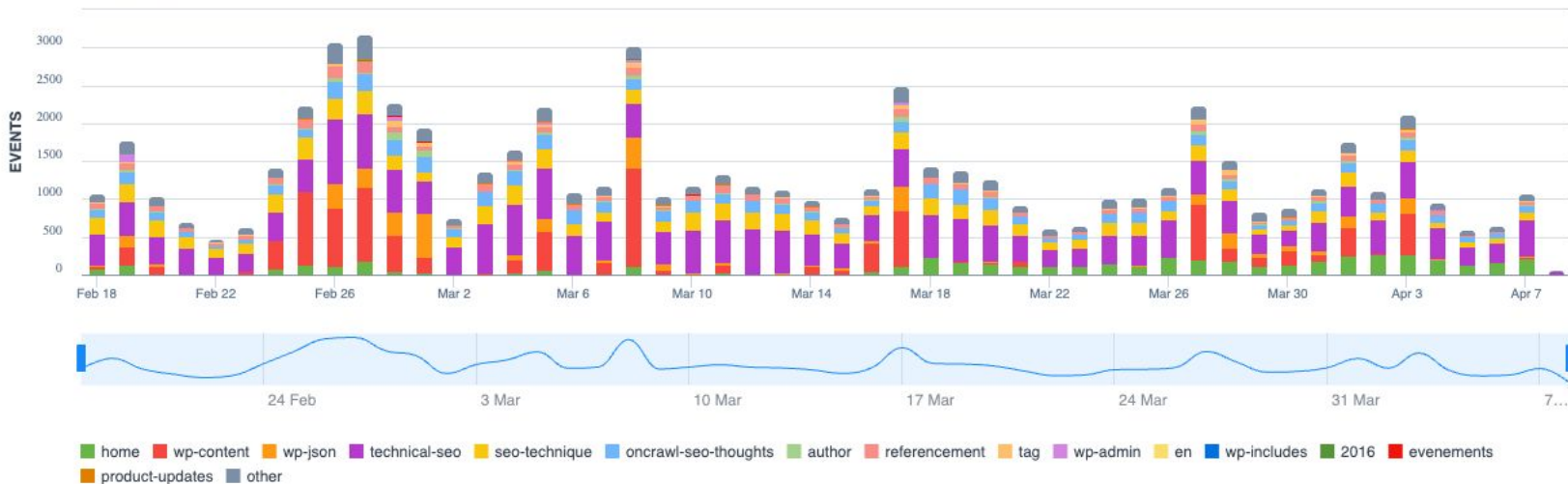
Identify accessibility issues

Status codes encountered by bots



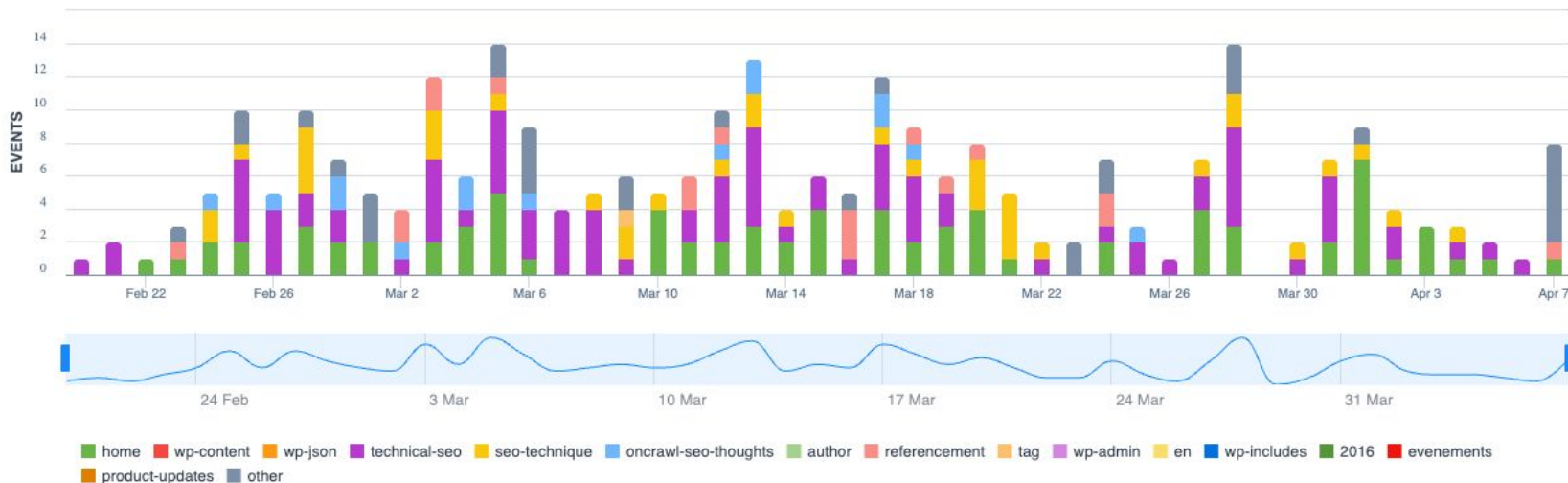
What types of pages are crawled by bots?

Bot hits by page group



What types of pages are generating clicks?

SEO visits



How can you use this data ?

- Make sure ChatGPT can access “useful” pages
- Identify pages crawled by ChatGPT that are not accessible
- Identify which type of content ChatGPT is crawling or not crawling
- Identify which type of content ChatGPT is crawling the most
- Identify the last time a page was crawled by ChatGPT

Segmentation recommendations

Here are some segmentations that will bring useful insights:

- Website page types
- Website categories
- Market / language
- Topic / taxonomy

Key Takeaways

Key takeaways

- **ChatGPT-user** and **OAI-SearchBot** are the key bots to monitor.
- You can block GPTBot but still appear in ChatGPT Search results.
- You can **use ChatGPT-user visits as a key visibility metric.**
- You can track **OAI-SearchBot to prevent indexation issues**
- Traditional technical SEO indexability / accessibility still matters.
- You can track your **ChatGPT traffic with logs.**
- Combining all this data will help you monitor your efforts and identify “ranking” patterns.

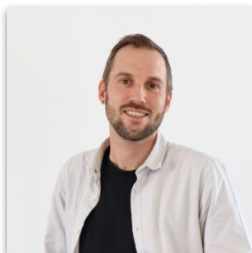
What's next?

Upcoming webinar

oncrawl × ROAST™

How to appear in ChatGPT: Practical SEO strategies for AI visibility

Speakers



Jérôme Salomon
Senior Technical SEO
@Oncrawl



John Campbell
Head of AI and Innovation
@ROAST

Save the date: May 06, 2025 – 3:00pm CET

Q&A

THANK YOU

www.oncrawl.com

